

Getting started with coding and R

stirlingcodingclub.github.io/getting_started/slides.pdf

Stirling Coding Club

20 October 2021

Stirling Coding Club

Stirling Coding Club exists for fun, informal meet-ups of friends and colleagues to share skills, stories, and ideas on using code, and exploring open research practices. Our goal is to create a friendly, no-pressure environment where people can share their work, ask for help on a coding problem, and learn and work together.

Stirling Coding Club

“Coding Club is for everyone, regardless of their career stage or current level of knowledge. Coding Club is a place that brings people together, regardless of their gender or background. We all have the right to learn, and we believe learning is more fun and efficient when we help each other along the way.” – [Edinburgh Coding Club](#)

Stirling Coding Club

- ▶ Alternates weeks with Stats Discussion Group
- ▶ Detailed notes posted on Monday
- ▶ Meeting on Wednesday, then try notes with live help or walk-through
- ▶ Questions on MS Teams chat afterwards
- ▶ Participate as much or as little as you prefer!

Stirling Coding Club in 2021-2022

1. Getting started in R (20 OCT 2021)
2. General computing concepts (03 NOV 2021)
3. Writing and using functions (17 NOV 2021)
4. Writing and using loops (01 DEC 2021)
5. Simulating data in R (12 JAN 2022)
6. Introduction to version control (26 JAN 2022)
7. Slides in Rmarkdown (09 FEB 2022)
8. Manuscripts in Rmarkdown (23 FEB 2022)
9. Rshiny applications (09 MAR 2022)
10. Randomisation in R (23 MAR 2022)
11. Writing tests for code (06 APR 2022)
12. Individual-based modelling (20 APR 2022)
13. Creating R packages (04 MAY 2022)

<https://stirlingcodingclub.github.io/studyGroup/>

Do I need to learn how to code?

Not only is coding a core skill that gets the basic work of biology done, it's also taught [biologists] to look at problems in new ways. Above all, they agree, **coding liberated them**. As tools evolve to allow biologists to gather ever-more-massive quantities of data, people [...] will find a way to make coding a core part of scientific education¹.

[T]here's nothing wrong with becoming an expert (or close to one) in a given skill set (e.g., statistics, coding), and **it also will be a good thing as a biologist**. But I don't think coding will be the must-have skill set a few years from now (except for those who develop new tools)².

¹Dreyfuss (2017). Wired.

<https://www.wired.com/2017/03/biologists-teaching-code-survive/>.

²Mike (2019). <https://mikethemadbiologist.com/2019/12/06/biologists-should-not-learn-how-to-code/>.

How does coding (specifically R) help me as a researcher?

Most scientific research relies on numeric data.

- ▶ Data organisation
- ▶ Statistical analysis
- ▶ Visualisation

How does coding (specifically R) help me as a researcher?

Most scientific research relies on numeric data.

- ▶ Data organisation
- ▶ Statistical analysis
- ▶ Visualisation

Software exists that can do the above without code.

- ▶ JMP
- ▶ Minitab
- ▶ Systat
- ▶ MSExcel

How does coding (specifically R) help me as a researcher?

Most scientific research relies on numeric data.

- ▶ Data organisation
- ▶ Statistical analysis
- ▶ Visualisation

Software exists that can do the above without code.

- ▶ JMP
- ▶ Minitab
- ▶ Systat
- ▶ MSExcel

Practical advantages of using these software tools

- ▶ Graphical User Interface (no need to work with code).

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source
- ▶ Thousands [R packages](#) for specific data needs

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source
- ▶ Thousands [R packages](#) for specific data needs
- ▶ Standard programming language for statistics

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source
- ▶ Thousands [R packages](#) for specific data needs
- ▶ Standard programming language for statistics
- ▶ Write [papers](#), [slides](#), and [apps](#) in Rmarkdown

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source
- ▶ Thousands [R packages](#) for specific data needs
- ▶ Standard programming language for statistics
- ▶ Write [papers](#), [slides](#), and [apps](#) in Rmarkdown
- ▶ **Write your own solutions to data management problems**
- ▶ **Write your own statistical analyses**
- ▶ **Create your own plots**

How does coding (specifically R) help me as a researcher?

Practical advantages of learning and using R

- ▶ R software is entirely free and open source
- ▶ Thousands [R packages](#) for specific data needs
- ▶ Standard programming language for statistics
- ▶ Write [papers](#), [slides](#), and [apps](#) in Rmarkdown
- ▶ **Write your own solutions to data management problems**
- ▶ **Write your own statistical analyses**
- ▶ **Create your own plots**

A lot of coding is Googling solutions to get the code to work.

Write your own solutions to data organisation problems

##		SPEI	Year	Tree_ID	BAI	cumul_mn
## 1	0.34325052	1966	FR6201	645.3972	NA	
## 2	-1.35933830	1967	FR6201	470.0363	NA	
## 3	0.49415034	1968	FR6201	830.5755	NA	
## 4	-1.38069918	1969	FR6201	414.0594	NA	
## 5	0.79613295	1970	FR6201	977.4877	NA	
## 6	-0.06371012	1971	FR6201	809.8834	NA	

Calc. the cumul_mn for the BAI col. every year for each tree¹:

¹Thanks to [Tom Ovenden](#) for letting me use this example.

Write your own solutions to data organisation problems

##		SPEI	Year	Tree_ID	BAI	cumul_mn
## 1	0.34325052	1966	FR6201	645.3972	NA	
## 2	-1.35933830	1967	FR6201	470.0363	NA	
## 3	0.49415034	1968	FR6201	830.5755	NA	
## 4	-1.38069918	1969	FR6201	414.0594	NA	
## 5	0.79613295	1970	FR6201	977.4877	NA	
## 6	-0.06371012	1971	FR6201	809.8834	NA	

Calc. the cumul_mn for the BAI col. every year for each tree¹:

- ▶ Do not include *current* BAI record in cumul_mn calc
- ▶ If the Year SPEI is >1 or <-1, never include the BAI value
- ▶ If the Year - 1 SPEI is >1 or <-1, never include the BAI value
- ▶ If the Year - 2 SPEI is >1.5 or <-1.5, never include the BAI value
- ▶ If the Year - 3 SPEI is >2 or <-2, never include the BAI value

¹Thanks to [Tom Ovenden](#) for letting me use this example.

Write your own solutions to data organisation problems

Solution took 65 lines of code written in 3 custom functions.

```
tree_cumulative_mean <- function(dat){  
  trees      <- unique(dat$Tree_ID);  
  new_table  <- NULL;  
  for(tree in trees){  
    sub_dat   <- dat[dat$Tree_ID == tree,];  
    tree_cumul <- get_cumulative(tree = sub_dat);  
    new_tree   <- update_cumul(tree = sub_dat,  
                               vec  = tree_cumul);  
    new_table  <- rbind(new_table, new_tree);  
  }  
  return(new_table);  
}
```

The above is the 'outermost' function.

Write your own statistical analyses

Testing the effects of fig tree spatial distributions on pollinator, seed, and parasite distributions.



¹Duthie AB, Nason JD (2016) *Oikos* [125\(11\):1597-1606](#).

Write your own statistical analyses

##	Year	Site	Tree	Poll	Npoll	Seeds	Vol	Lat
## 1	2010	170	6	0	22	NA	753.6000	28.24057
## 2	2010	170	6	0	10	NA	1025.7333	28.24057
## 3	2010	170	6	7	37	NA	1047.7133	28.24057
## 4	2010	170	6	0	36	NA	1230.8800	28.24057
## 5	2010	170	6	0	34	NA	884.4333	28.24057
## 6	2010	170	6	0	33	NA	1128.3067	28.24057
## 7	2010	170	6	0	36	NA	1025.7333	28.24057
## 8	2010	170	6	0	17	NA	565.2000	28.24057

Write your own statistical analyses

##	Year	Site	Tree	Poll	Npoll	Seeds	Vol	Lat
## 1	2010	170	6	0	22	NA	753.6000	28.24057
## 2	2010	170	6	0	10	NA	1025.7333	28.24057
## 3	2010	170	6	7	37	NA	1047.7133	28.24057
## 4	2010	170	6	0	36	NA	1230.8800	28.24057
## 5	2010	170	6	0	34	NA	884.4333	28.24057
## 6	2010	170	6	0	33	NA	1128.3067	28.24057
## 7	2010	170	6	0	36	NA	1025.7333	28.24057
## 8	2010	170	6	0	17	NA	565.2000	28.24057

- ▶ Needed to use mean tree values and a custom index of fig wasp dispersal ability

Write your own statistical analyses

##	Year	Site	Tree	Poll	Npoll	Seeds	Vol	Lat
## 1	2010	170	6	0	22	NA	753.6000	28.24057
## 2	2010	170	6	0	10	NA	1025.7333	28.24057
## 3	2010	170	6	7	37	NA	1047.7133	28.24057
## 4	2010	170	6	0	36	NA	1230.8800	28.24057
## 5	2010	170	6	0	34	NA	884.4333	28.24057
## 6	2010	170	6	0	33	NA	1128.3067	28.24057
## 7	2010	170	6	0	36	NA	1025.7333	28.24057
## 8	2010	170	6	0	17	NA	565.2000	28.24057

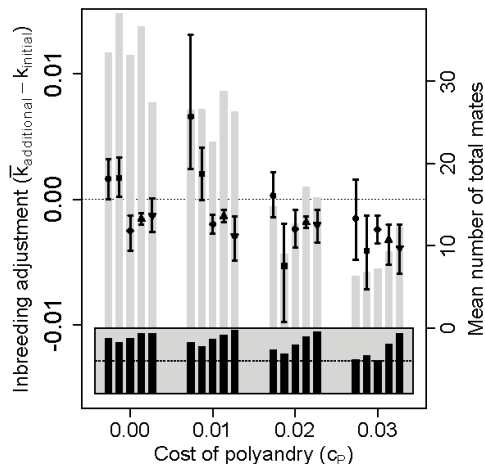
- ▶ Needed to use mean tree values and a custom index of fig wasp dispersal ability
- ▶ Wanted to make [analysis open source](#) and figures reproducible

Write your own statistical analyses

##	Year	Site	Tree	Poll	Npoll	Seeds	Vol	Lat
## 1	2010	170	6	0	22	NA	753.6000	28.24057
## 2	2010	170	6	0	10	NA	1025.7333	28.24057
## 3	2010	170	6	7	37	NA	1047.7133	28.24057
## 4	2010	170	6	0	36	NA	1230.8800	28.24057
## 5	2010	170	6	0	34	NA	884.4333	28.24057
## 6	2010	170	6	0	33	NA	1128.3067	28.24057
## 7	2010	170	6	0	36	NA	1025.7333	28.24057
## 8	2010	170	6	0	17	NA	565.2000	28.24057

- ▶ Needed to use mean tree values and a custom index of fig wasp dispersal ability
- ▶ Wanted to make [analysis open source](#) and figures reproducible
- ▶ Had to help a colleague use the same analysis on new data

Create your own plots



Custom built plot [with R code](#) showing results from an individual-based model.

¹Duthie AB, et al. (2016) *Evolution* [70\(9\)](#), 1927–1943.

R is extremely useful for doing scientific research

A bit of coding skill can save a lot of time in the long run.

- ▶ You do not need to know everything
- ▶ Help is available from many places



https://stirlingcodingclub.github.io/getting_started/notes.html

<https://ourcodingclub.github.io/tutorials/intro-to-r/>

<https://ourcodingclub.github.io/tutorials/troubleshooting/>